We are IntechOpen, the world’s leading publisher of Open Access books
Built by scientists, for scientists

3,800
Open access books available

116,000
International authors and editors

120M
Downloads

154
Countries delivered to

TOP 1%
Our authors are among the most cited scientists

12.2%
Contributors from top 500 universities

WEB OF SCIENCE™
Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com
1. Introduction

An outlier is an observation (or measurement) that is different with respect to the other values contained in a given dataset. Outliers can be due to several causes. The measurement can be incorrectly observed, recorded or entered into the process computer, the observed datum can come from a different population with respect to the normal situation and thus is correctly measured but represents a rare event. In literature different definitions of outlier exist: the most commonly referred are reported in the following:

- “An outlier is an observation that deviates so much from other observations as to arouse suspicions that is was generated by a different mechanism” (Hawkins, 1980).
- “An outlier is an observation (or subset of observations) which appear to be inconsistent with the remainder of the dataset” (Barnet & Lewis, 1994).
- “An outlier is an observation that lies outside the overall pattern of a distribution” (Moore and McCabe, 1999).
- “Outliers are those data records that do not follow any pattern in an application” (Chen and al., 2002).
- “An outlier in a set of data is an observation or a point that is considerably dissimilar or inconsistent with the remainder of the data” (Ramasmawy at al., 2000).

Many data mining algorithms try to minimize the influence of outliers for instance on a final model to develop, or to eliminate them in the data pre-processing phase. However, a data miner should be careful when automatically detecting and eliminating outliers because, if the data are correct, their elimination can cause the loss of important hidden information (Kantardzic, 2003). Some data mining applications are focused on outlier detection and they are the essential result of a data-analysis (Sane & Ghatol, 2006).

The outlier detection techniques find applications in credit card fraud, network robustness analysis, network intrusion detection, financial applications and marketing (Han & Kamber, 2001). A more exhaustive list of applications that exploit outlier detection is provided below (Hodge, 2004):

- Fraud detection: fraudulent applications for credit cards, state benefits or fraudulent usage of credit cards or mobile phones.
- Loan application processing: fraudulent applications or potentially problematical customers.
- Intrusion detection, such as unauthorized access in computer networks.
- Activity monitoring: for instance the detection of mobile phone fraud by monitoring phone activity or suspicious trades in the equity markets.
- Network performance: monitoring of the performance of computer networks, for example to detect network bottlenecks.
- Fault diagnosis: processes monitoring to detect faults for instance in motors, generators, pipelines.
- Structural defect detection, such as monitoring of manufacturing lines to detect faulty production runs.
- Satellite image analysis: identification of novel features or misclassified features.
- Detecting novelties in images (for robot neotaxis or surveillance systems).
- Motion segmentation: such as detection of the features of moving images independently on the background.
- Time-series monitoring: monitoring of safety critical applications such as drilling or high-speed milling.
- Medical condition monitoring (such as heart rate monitors).
- Pharmaceutical research (identifying novel molecular structures).
- Detecting novelty in text. To detect the onset of news stories, for topic detection and tracking or for traders to pinpoint equity, commodities.
- Detecting unexpected entries in databases (in data mining application, to the aim of detecting errors, frauds or valid but unexpected entries).
- Detecting mislabeled data in a training data set.

How the outlier detection system deals with the outlier depends on the application area. A system should use a classification algorithm that is robust to outliers to model data with naturally occurring outlier points. In any case the system must detect outlier in real time and alert the system administrator. Once the situation has been handled, the anomalous reading may be separately stored for comparison with any new case but would probably not be stored with the main system data as these techniques tend to model normality and use outliers to detect anomalies (Hodge, 2004).

2. Traditional approaches

The salient traditional approaches to outlier detection can be classified as either distribution-based, depth-based, clustering, distance-based or density-based.

2.1 Distribution-based method

These methods are typically found in statistics textbooks. They deploy some standard distribution model (Normal, Poisson, etc.) and flag as outliers those data which deviate from the model. However, most distribution models typically apply directly to the future space and are univariate i.e. have very few degrees of freedom. Thus, they are unsuitable even for moderately high-dimensional data sets. Furthermore, for arbitrary data sets without any prior knowledge of the distribution of points, expensive tests are required to determine which model best fits the data, if any! (Papadimitriou et al., 2002) Fitting the data with standard distributions is costly and may not produce satisfactory results. The most popular distribution is the Gaussian function. (Kim & Cho, 2006).

A method was proposed by Grubbs which calculates a Z value as the difference between the mean value for the attribute and the query value divided by the standard deviation for the


The book presents an excellent overview of the recent developments in the different areas of Robotics, Automation and Control. Through its 24 chapters, this book presents topics related to control and robot design; it also introduces new mathematical tools and techniques devoted to improve the system modeling and control. An important point is the use of rational agents and heuristic techniques to cope with the computational complexity required for controlling complex systems. Through this book, we also find navigation and vision algorithms, automatic handwritten comprehension and speech recognition systems that will be included in the next generation of productive systems developed by man.

How to reference
In order to correctly reference this scholarly work, feel free to copy and paste the following:
